Complaint Classification using Neural Networks and NLP

Objective:

Build a model that can classify customer complaints into predefined categories using Natural Language Processing (NLP) and Neural Networks.

Project Tasks:

1. Data Understanding & Preprocessing

- Perform **Exploratory Data Analysis (EDA)**:
  - Class distribution
  - Complaint length analysis
  - Word cloud for each category (optional)
- Text Cleaning:
  - Lowercasing
  - Removing punctuation
  - Removing stopwords
  - Lemmatization or stemming

2. Text Vectorization (Feature Extraction)

- **Mandatory**: Use **Word Embeddings**:
  - Pre-trained: GloVe, Word2Vec, or FastText (recommended)
  - Or train embeddings from scratch using Embedding Layer in Keras/PyTorch
- Optional: Compare with TF-IDF + Dense Neural Networks

3. Model Building (Neural Networks)

- **Baseline**: Build a simple feedforward Neural Network using embeddings.
- **Advanced**: Build a Recurrent Neural Network (RNN) or LSTM model for sequence modelling.
- **Optional**: Try ID Convolutional Neural Networks (CNNs) for text classification.

4. Model Evaluation

- Accuracy, Precision, Recall, Fl-Score per class
- Confusion Matrix
- Training vs. Validation loss and accuracy plots

5. Model Optimization

- Try different optimizers, batch sizes, learning rates.
- Experiment with dropout, regularization, embedding size.

Rules:

- Do not use traditional ML models (like SVM, Random Forest, etc.).
- Must use neural networks (Feedforward, LSTM, GRU, CNN etc.).
- No direct copying of code from online tutorials; they can take references but the implementation must be custom.

Bonus Ideas:

- Visualize attention weights (if using advanced models)
- Handle multi-label classification if the dataset allows
- Perform error analysis on misclassified samples